Enhancing Modality-Agnostic Representations via Meta-learning for Brain Tumor Segmentation

Aishik Konwer[1], Xiaoling Hu[1], Joseph Bae[2], Xuan Xu[1], Chao Chen[2], Prateek Prasanna[2]
[1]Department of Computer Science, Stony Brook University
[2]Department of Biomedical Informatics, Stony Brook University
{akonwer,xiaolhu,xuaxu}@cs.stonybrook.edu
{joseph.bae,chao.chen.1,prateek.prasanna}@stonybrook.edu

Abstract

In medical vision, different imaging modalities provide complementary information. However, in practice, not all modalities may be available during inference or even training. Previous approaches, e.g., knowledge distillation or image synthesis, often assume the availability of full modalities for all subjects during training; this is unrealistic and impractical due to the variability in data collection across sites. We propose a novel approach to learn enhanced modality-agnostic representations by employing a meta-learning strategy in training, even when only limited full modality samples are available. Meta-learning enhances partial modality representations to full modality representations by meta-training on partial modality data and meta-testing on limited full modality samples. Additionally, we co-supervise this feature enrichment by introducing an auxiliary adversarial learning branch. More specifically, a missing modality detector is used as a discriminator to mimic the full modality setting. Our segmentation framework significantly outperforms state-of-the-art brain tumor segmentation techniques in missing modality scenarios.

1. Introduction

Multiple medical imaging modalities/protocols are required to provide complementary diagnostic cues to clinicians. For instance, multiple Magnetic Resonance Imaging (MRI) sequences (henceforth referred to as modalities), namely native T1, post-contrast T1 (T1c), T2-weighted (T2), and Fluid Attenuated Inversion Recovery (FLAIR) are used together to understand the underlying spatial complexity of brain tumors and their surroundings [3, 5]. Deep learning approaches [21, 36, 10, 58, 51, 49] have found great success in multimodal brain tumor segmentation and treatment response assessment. These conventional brain tumor segmentation methods perform well only when all

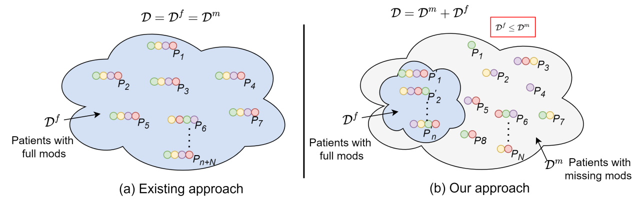


Figure 1: Comparison of the paradigms generally adopted by existing missing modality approaches (left) vs. ours (right) for brain tumor segmentation. N and n refer to the number of subjects (patients) with partial and full modalities, respectively. Previous methods either utilize full modality data $\mathcal{D}^f$ for all subjects or simulate partial modality data $\mathcal{D}^m$ from $\mathcal{D}^f$. On the contrary, our approach works in a limited full modality setting, i.e., $|\mathcal{D}^f| \leq |\mathcal{D}^m|$.

four acquisition modalities are available as input (i.e., in the *full modality setting*). However, in clinical practice, often only a subset of modalities are available due to issues including image degradation, motion artifacts [18], erroneous acquisition settings, and brief scan times. Hence it is crucial to develop robust modality-agnostic methods which can achieve state-of-the-art performance in *missing modality settings*, i.e., when different modalities are unavailable during inference or even training.

Recently, a plethora of works has been proposed to address missing modality scenarios for brain tumor segmentation. Two major categories include: 1) Knowledge distillation: These methods [43, 23, 52, 50, 2] learn privileged information from a teacher network trained on full modality data, i.e., data with all modalities available. 2) Image synthesis: Several works [42, 54, 59, 25, 55] train generative models to synthesize images of the missing modalities. The synthesized "full modality" images are used for segmentation. One major issue is that both categories of methods require full modality data for all subjects in the training set (see Fig. 1a), either to train the teacher or the generator. This can be very unrealistic; in real-world applications, most studies only have very limited full modality data, far from sufficient for training. In this paper, we focus on a more realistic setting: most training data is only partial modality data, i.e., having a few modalities missing. We ask the following question: *How do we efficiently learn from a large amount of partial modality data and a small amount of full modality data (see Fig. 1b)?*

Another category recently rising in popularity is Shared Latent Space modeling [22, 14, 28, 7, 13, 56, 57, 53, 20, 30, 60, 8]. These methods learn a shared latent representation from partial modality data. However, the quality of the learned representation can be limited by the heterogeneity of available modalities. The learned representation will be biased towards the most frequently available modalities and essentially overlook minority modalities (i.e., modalities that appear less frequently in training). This will inevitably lead to sub-optimal performance on test data, especially with minority modalities. To compensate for this undesirable bias, these methods often resort to segmenting all modalities individually from the shared representation, ultimately requiring full modality for all cases during training.

These observations, further summarized in Tab. 1, motivate us to design a modality-agnostic method that can fully utilize partial modality data. Through the usage of the meta-learning strategy, our method learns enriched shared representations that are generalizable and not biased towards more frequent modalities, even with limited full modality data.

Category	Can handle limited FM?	Learns Unbiased mapping?
KD [23, 52, 50, 2]	N	Y
GAN [42, 54, 59, 25, 55]	N	Y
Shared (others) [8, 13, 56]	N	N
SMIL [31]	Y	N
Shared (Ours)	Y	Y

Table 1: Advantages of our approach over existing frameworks. We are able to train in a limited full modality (FM) setting (with $\leq 50\%$ FM samples), and learn an unbiased mapping that is unaffected by the proportion of any given modality in training.

Our core idea is based on the meta-learning technique [16]. Meta-learning provides an effective framework to learn to perform multiple tasks in a mutually beneficial manner. We consider segmentation with each partial modality input combination as a different task, yielding $2^M - 1$ meta-tasks for M modalities. By learning all meta-tasks in parallel, meta-learning ensures the network generates modality-agnostic representations. Thanks to meta-learning, tasks depending on rare modalities can be significantly improved even with limited training data. This maximally mitigates the bias against rare modalities. Meanwhile, we propose using a small amount of full modality

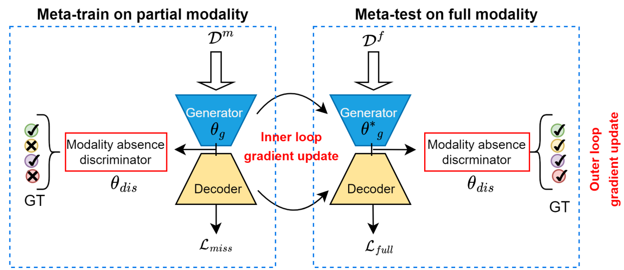


Figure 2: Framework overview. $\mathcal{D}^m$ (partial modality) and $\mathcal{D}^f$ (full modality) are used as inputs for encoder-decoder networks in the meta-train and meta-test phase, respectively. Partial modality representations are adapted to the full modality domain via: 1) meta-optimization of gradients in both data, and 2) adversarial learning based on predictions by a modality absence classifier.

data only during meta-testing. Meta-testing is introduced as an intermediate step in meta-learning to boost the generalization performance of the model across different tasks. Using full modality data, albeit limited, in meta-test can maximally leverage such data to enhance the representation quality of the model. This innovative meta-learning design ensures we learn with a large amount of partial modality data and only a small amount of full modality data, with negligible partial modality bias.

Recently a meta-learning approach [31] performed classification with missing modalities. They predict the prior weights of modalities via a feature reconstruction network, the quality of which is indirectly dependent on the number of full modality samples. This method is unsuitable for our segmentation framework since conventional approaches (PCA [40], K-Means [33]) cannot be used to cluster the priors in a high dimensional latent space. Moreover, [31] deals with only two input modalities and considers them individually as meta-tasks, while we construct a heterogeneous task distribution with different combinations of inputs respecting the heterogeneity settings of real-world data.

We also employ a novel adversarial learning technique that further enhances the quality of the generated shared latent space representation. Previous GAN-based approaches [42] reconstruct the missing modalities in image space; this leads to the impractical requirement of full modality as ground truth for training. Our task is achieved in latent space by designing the discriminator as a multi-label classifier. The discriminator predicts the presence/absence of modalities from the fused latent representation performing a binary classification for each modality. Our ultimate goal is to hallucinate the full modality representation from the hetero-modal feature space. Note that due to the hetero-modal nature of the data, the number of available modalities can vary dramatically across subjects. To address this, we utilized a channel-attention weighted fusion module that can accept a varying number of representations as input but generates a single fused output.

Overall, our contributions can be summarized as follows:

- We propose a meta-learning paradigm to train with hybrid data (partial and full modalities) and also enhance the learned partial modality representations to mimic a full modality representation. This is accomplished by meta-training on partial modality data while finetuning on limited full modality data during the meta-test. Such a training strategy overcomes the over-reliance on full modality data, as well as succeeds in learning an unbiased representation for all missing situations.
- We introduce a novel adversarial learning strategy to further enrich the shared representations in the latent space. It differs from other generative approaches that synthesize missing images and demand full modality ground truths for training. Our approach does not necessitate reconstructing missing modality images.

2. Related Work

Segmentation with missing modalities. Incomplete data is a long-standing issue in computer vision; it particularly has significant implications in medical vision. Due to privacy concerns and budget constraints, one or more modalities (audio/visual/text) [27, 15, 11, 48, 6] of a given sample may not be available. In this work, we focus on partial medical imaging modalities for brain tumor segmentation. Existing methods on complete multi-modal brain tumor segmentation [21, 36, 10, 58, 51, 49] perform poorly in realistic hetero-modal settings.

Researchers have broadly used three techniques to perform brain tumor segmentation from missing modalities including *knowledge distillation* (KD) [43, 23, 52, 50, 2], *generative modeling* [42, 54, 59, 25, 55] and *shared representation learning* [22, 14, 28, 7, 13, 56, 57, 53, 20, 30, 60, 8]. ACN [52] trains a separate teacher-student pipeline for each subset of modalities. Among the generative models, MM-GAN [42] uses a U-Net to impute missing modalities while a PatchGAN learns to discriminate between real and synthesized inputs. A major drawback of KD and GAN-based approaches is their inability to perform well when all modalities are not present for a subject during training. Moreover, the unstable and non-converging nature of a 3D generator may lead to degraded quality of synthesized images, eventually affecting downstream performance. Our method belongs to the third category, i.e., shared latent space models. In [22, 14], the authors compute variational statistics to construct unified representations for segmentation. Multi-source information is modeled using a correlation constraint [60] or region-aware fusion blocks [13] to encode shared representations. Recent frameworks [57, 56] in this genre advocate for exploiting intra/inter-modality relations through graph and transformer-based modeling. Such approaches usually lack flexibility for adaptation to all missing scenarios. They yield sub-optimal performance due to failure in the retrieval of discriminative features generally existing in full modality data. Furthermore, these approaches can learn biased mappings among the available modalities leading to poor generalizability for modalities not encountered in training.

Meta-learning. Meta-learning algorithms [16, 37, 47] are inspired by human perception of new tasks. Optimization-based meta-learning techniques [16, 37, 1] have gained popularity since they can easily encode prior information through an optimization process. Model-agnostic meta-learning (MAML) [45] is the most commonly used algorithm under this category, due to its flexible application to any network trained through gradient descent. Researchers have widely adopted MAML frameworks to generalize a model to new tasks, unseen domains, and enriching input features in multimodal scenarios [29, 32]. SMIL [31] introduces a Bayesian MAML framework that attains comparable classification performance across both partial and full modality data. However, their approach requires prior reconstruction of the missing modalities. HetMAML [9] can handle heterogeneous task distributions, i.e. different modality combinations for input space but fails to attain generalizable performance across partial and full modalities. Inspired by the above two approaches, we propose a modality-agnostic architecture that can not only accept hetero-modal inputs but also enhance their representations with the additional information present in a full set of modalities. This leads to better segmentation performance for any hetero-modal input instance.

Domain adaptation. Domain adaptation refers to the training of a neural network to jointly generate both discriminative and domain-invariant features in order to model different source and target data distributions [17, 4, 19, 46, 26]. Authors in [17] leverage an auxiliary domain classifier to address the domain shift. Inspired by this approach, we design our discriminator as a modality absence predictor. Similar to Sharma et al. [42], we feed our discriminator with the correct modality code as ground truth, while the generator is provided an 'all-one' full modality code impersonating the presence of all modalities. In an attempt to fool the discriminator, the generator learns to always mimic full modality representations, irrespective of the available inputs. This results in enhanced representations that boost downstream performance in missing modality situations.

3. Methodology

Overview. Given heterogeneous modalities as input, our goal is to build a modality-agnostic framework that can be robust to missing modality scenarios, and achieve performances comparable to a full modality setting. We have lim-

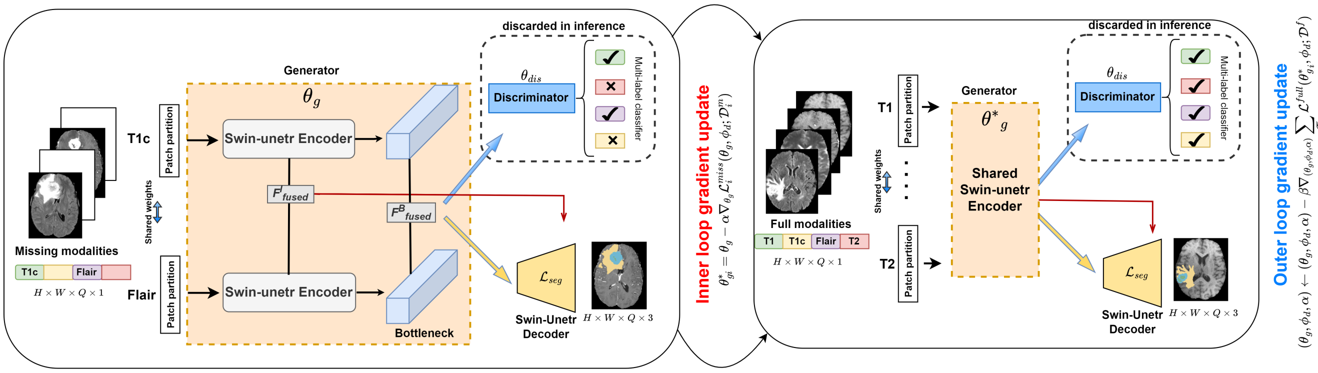


Figure 3: Illustration of the proposed framework. Available or full set of modalities are passed through a shared generator in the meta-train and meta-test stages respectively. The aggregation module helps to obtain a fused representation from five different levels (level l and bottleneck are indicated here). Next, only the bottleneck embedding is used by the discriminator to predict which modalities are present at the input. All five fused embeddings are used by the segmentation decoder. Inner and outer loop gradient updates refer to the losses calculated in the meta-train and meta-test stages on partial modality and full modality data, respectively.

ited access to full modality data during training; this simulates a practical clinical scenario where brain tumor segmentation may need to be performed with partial modalities. To address this data-scarce situation, we aim to establish a mapping between partial and full modality representations. Our proposed approach is shown in Fig. 3. Meta-learning has been shown to be an efficient computational paradigm in dealing with heterogeneous training data [9], or conducting feature adaptation between different domains [31]. To this end, we adapt model-agnostic meta-learning to leverage information from both partial and full modality data. This strategy is elaborated in Sec. 3.1. We also want to further enrich encoded representations obtained from available modalities, with the supplemental information contained in full modality representations. More specifically, we propose a novel adversarial learning technique introducing a discriminator that acts as a modality absence classifier. A detailed description is provided in Sec. 3.2. Because we need to generate a common fused representation for each hetero-modal input combination, our architecture incorporates a simple and elegant feature aggregation module (see Sec. 3.3).

3.1. Meta-Learning for Feature Adaptation

Suppose we have a total of M MRI modalities as inputs for a patient. To simulate a real-world clinical scenario where full modalities may only be available for a fraction of subjects, some modalities are dropped for each patient during training. This training paradigm ensures the model becomes more robust to missing scenarios at inference. Thus we construct a heterogeneous task distribution $P(\mathcal{T})$ that is a collection of k task distributions: $P(\mathcal{T}^1), P(\mathcal{T}^2), ..., P(\mathcal{T}^k)$. Each such distribution $P(\mathcal{T}^i)$ has a distinct input feature space related to a specific subset of modalities. We however exclude the full modality subset from the task distribution due to its utilization in meta-testing, as explained in the following paragraph. Overall, k types of task instances can be sampled from $P(\mathcal{T})$, where $k = 2^M - 2$.

Algorithm 1 Modality-Agnostic Meta-Learning

Input: Training dataset $\mathcal{D}$ is divided into two cohorts of subjects with partial/missing and full modalities respectively $\mathcal{D} = \{\mathcal{D}^m, \mathcal{D}^f\}$; β is the learning rate.
Initialise: Initialise $\theta_g, \phi_d = \{\theta_{dis}, \theta_{dec}\}, \alpha$
Output: Optimized meta-parameters $\{\theta_g, \phi_d, \alpha\}$
while not converged **do**
    Sample a batch of tasks $\mathcal{T}^i \sim \{P(\mathcal{T})\}$
    for each task $\mathcal{T}^i$ **do**
        Evaluate inner loop loss: $\mathcal{L}^{miss}_i(\theta_g, \phi_d; \mathcal{D}^m_i)$
        Adapt: $\theta^*_{gi} = \theta_g - \alpha\nabla_{\theta_g}\mathcal{L}^{miss}_i(\theta_g, \phi_d; \mathcal{D}^m_i)$
        Compute outer loop loss: $\mathcal{L}^{full}(\theta^*_{g\,i}, \phi_d; \mathcal{D}^f)$
    end for
    Update meta-parameters: $(\theta_g, \phi_d, \alpha) \leftarrow (\theta_g, \phi_d, \alpha) - \beta\nabla_{(\theta_g, \phi_d, \alpha)} \sum_{\mathcal{T}_i} \mathcal{L}^{full}(\theta^*_{g\,i}, \phi_d; \mathcal{D}^f)$
end while

Formally, we have a hetero-modal training dataset $\mathcal{D}$ which we divide into two cohorts of subjects $\{\mathcal{D}^m, \mathcal{D}^f\}$ containing partial and full modalities, respectively. The goal is to effectively learn from both types of data. We construct a batch of subjects $\mathcal{D}^m_i$ corresponding to each $P(\mathcal{T}^i)$. The pair of subjects and their corresponding task remains fixed over all epochs. Only the modalities which are not included in a task get dropped for that particular subject.

Shared encoders are used along with a fusion module to produce a modality-agnostic representation. In our case, both encoder and fusion modules jointly constitute the generator E_{θ_g} (parameterized by θ_g). An MLP-based classifier network, parameterized by θ_{dis}, is employed as a discriminator as explained in Sec. 3.2. For clarity, parameters of the discriminator and decoder network, $\{\theta_{dis}, \theta_{dec}\}$, are collectively symbolized as ϕ_d. Our aim is to obtain an optimal generator parameter θ_g through task-wise training on $\mathcal{D}_i^m$ by reducing the inner loop objective $\mathcal{L}_i^{miss}$.

$$\theta_{gi}^* = \theta_g - \alpha \nabla_{\theta_g} \mathcal{L}_i^{miss}(\theta_g, \phi_d; \mathcal{D}_i^m), \tag{1}$$

where α is a learnable rate for inner-level optimization. The optimized model is expected to perform better on $\mathcal{D}^f$. The goal of the updated framework is to accomplish the outer loop objective $\mathcal{L}^{full}$ across all sampled tasks:

$$\min_{\theta_g, \phi_d} \sum_{\mathcal{T}_i} \mathcal{L}^{full}(\theta_{g_i}^*, \phi_d; \mathcal{D}^f). \tag{2}$$

Both the inner and outer loop losses are kept the same, referring to the generator and discriminator losses, $\mathcal{L}_E$ and $\mathcal{L}_{dis}$. By forcing the partial modality trained model to perform well on full modality data, we implicitly target the recovery of relevant information for better segmentation in missing modality scenarios. This partial to full modality mapping in feature space, is further strengthened by the introduction of a domain-adaptation inspired feature enrichment module (Details in Sec. 3.2). All three meta parameters $(\theta_g, \phi_d, \alpha)$ are henceforth meta-updated by averaging gradients of outer loop loss over a meta-batch of tasks.

$$(\theta_g, \phi_d, \alpha) \leftarrow (\theta_g, \phi_d, \alpha) - \beta \nabla_{(\theta_g, \phi_d, \alpha)} \sum_{\mathcal{T}_i} \mathcal{L}^{full}(\theta_{g_i}^*, \phi_d; \mathcal{D}^f). \tag{3}$$

Thus during meta-training, the model tunes its initialization parameter to achieve improved generalizability across all missing modality tasks. During meta-test, by fine-tuning with full modality data, we map the learned feature representations to the full modality space. Different from MAML, the pretrained model is directly evaluated on datasets where subjects contain a fixed subset of modalities (one of the tasks $\mathcal{T}^i$ already encountered in meta-training) at inference. The training process is summarized in Alg. 1.

3.2. Adversarial Feature Enrichment

Considering that full modality data contains richer information, we enforce encoder outputs to mimic full representations, irrespective of the limited input combination. The modality encoders and fusion module can be collectively considered as a shared generator E. We introduce an MLP-based multi-label classifier as our discriminator D.

The objective of D is to predict the absence/presence of modalities from the fused embedding $\mathbf{F}_{fused}^B$ at the bottleneck level. D utilizes Binary Cross-Entropy loss $\mathcal{L}_{\mathcal{BCE}}$, and sigmoid activation to output M binary predictions $\hat{d}$, denoting whether a modality is available or not. While calculating the discriminator loss $\mathcal{L}_{dis}$ indicated below, the ground truth variable T_{real} is a vector of size M which reflects the true combination of modalities available at input for that iteration. For example, assuming that $M = 4$, and only first two modalities are available, $T_{real} = \{1, 1, 0, 0\}$.

$$\mathcal{L}_{dis} = \sum_{z=1}^{\mathcal{D}^m + \mathcal{D}^f} \mathcal{L}_{\mathcal{BCE}}(\hat{d}_z, T_{real_z}). \tag{4}$$

The generator loss is a combination of segmentation loss and an adversarial loss used to train the generator to fool the discriminator. We consider a dummy ground truth variable T_{dummy}. In order to encourage the generator to encode representations that confuse or "fool" the discriminator into inferring that all modalities are present, we set $T_{dummy} = \{1, 1, 1, 1\}$, masquerading all generated representations as full modality representations. Thus D pushes the generator E to agnostically produce full modality representations.

$$\mathcal{L}_E = \lambda_1 \mathcal{L}_{seg} + \lambda_2 \sum_{z=1}^{\mathcal{D}^m + \mathcal{D}^f} \mathcal{L}_{\mathcal{BCE}}(\hat{d}_z, T_{dummy_z}). \tag{5}$$

3.3. Modality-Agnostic Feature Aggregation

We aim to utilize multiple modalities (which vary in number per patient) and derive a common fused representation. Individual encoders $E_1, E_2, ..., E_n$ having shared parameters are trained to extract features from each of the n available patient-specific modalities, where $1 \leq n \leq M$. These features $\mathbf{F}_1^l, \mathbf{F}_2^l, ..., \mathbf{F}_n^l$ obtained from the corresponding levels (l) of each encoder are passed into a feature aggregation module.

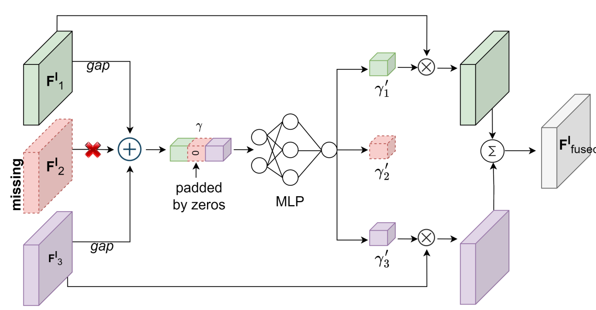


Figure 4: Illustration of the feature aggregation module. Modality $\mathbf{F}_2^l$ is missing. $\mathbf{F}_1^l$ and $\mathbf{F}_3^l$ are passed through global average pooling (GAP) operation and eventually fed into an MLP to generate the shared representation $\mathbf{F}_{fused}^l$.

The individual encoded representations undergo a Global Average Pooling operation and are subsequently concatenated to form a M-dimensional vector γ. This is achieved by imputing zeros in the channel information of $(M-n)$ missing modalities. γ is mapped to the channel weights of M modality features through a multi-layer perceptron (MLP) and sigmoid activation function, σ. These modality-specific weights multiplied with the corresponding features give rise to the aggregated representation, $\mathbf{F}^{l}_{fused}$ (in Fig. 4), which is eventually used as input to the decoder for segmentation. Our aggregation module exploits the correlation among available modality representations to create a unified feature that best describes the tumor characteristics of a subject. Detailed explanations with equations regarding this module can be found in the supplementary (Sec. 12).

We adopt a Swin-UNETR [49] architecture that employs soft Dice loss [35] to perform voxel-wise semantic segmentation. The segmentation loss function $\mathcal{L}_{Seg}$ is defined as follows:

$$\mathcal{L}_{seg}(G,P) = 1 - \frac{2}{V}\sum_{v=1}^{V}\frac{\sum_{u=1}^{U} G_{u,v}P_{u,v}}{\sum_{u=1}^{U} G_{u,v}^{2} + \sum_{u=1}^{U} P_{u,v}^{2}}.$$

where V is the number of classes and U is the number of voxels. $P_{u,v}$ and $G_{u,v}$ refer to the predicted output and one-hot encoded ground truth for class v at voxel u, respectively.

4. Experiment Design and Results

To validate our framework in various missing scenarios, we evaluate brain tumor segmentation results on all fifteen combinations of the four image modalities for a fixed test set. The average score is also reported for comparisons.

Datasets. We use three segmentation datasets from BRATS2018, BRATS2019, and BRATS2020 challenges [34]. They comprise 285, 335, and 369 training cases respectively. All subjects have $M = 4$ MR sequences. We perform 3D volumetric segmentation with images of size $155 \times 240 \times 240$. Pre-processing details are contained in the supplementary (Sec. 15). The segmentation classes include whole tumor (WT), tumor core (TC), and enhancing tumor (ET). Additional segmentation results on two non-BRATS hetero-modal cohorts (with different medical imaging modalities such as CT and MRI) are also reported in the supplementary (Sec. 16).

Implementation details. The experiments are implemented in Pytorch 1.7 [39] with three 48 GB Nvidia Quadro RTX 8000 GPUs. We drop modalities during training to construct the missing-modality dataset $\mathcal{D}^m$. We randomly sample a set of subjects from $\mathcal{D}^m_i$ assigned to each task distribution $P(\mathcal{T}^i)$. For a given subject, only those modalities which are not present in its associated task distribution are dropped. We ablate the fraction of subjects reserved for the full modality dataset $\mathcal{D}^f$ (Fig. 7a). Our method adopts a modified Swin-UNETR [49] housing up to 4 encoders E_1, E_2, E_3, E_4 which are Swin transformers (see supplementary Sec. 12). Both MLPs for the discriminator and feature aggregation are fully connected networks whose hidden layer dimensions are 48 and 64 respectively. The images are first resized to $128 \times 128 \times 128$, which is kept consistent across all compared methods. Features are extracted from 5 different levels of each encoder. For training and testing cohorts, we randomly split BRATS2018 into 200 and 85 subjects, BRATS2019 into 250 and 85 subjects, and BRATS2020 into 269 and 100 subjects, respectively. The batch size per task is kept as 1. During meta-training, we consider a metabatch size of 8, i.e., our meta-batch comprises 8 different modality combinations, each representing a separate task $\mathcal{T}_i$. More details can be found in the supplementary (Sec. 15). During inference, the meta-pretrained model is evaluated on test sets where all subjects have a fixed subset of modalities. The discriminator is discarded at inference.

Performance metrics. Dice similarity coefficient (DSC $\uparrow$) (Tab. 2) and Hausdorff Distance (HD95 $\downarrow$) (see supplementary Sec. 9) are used to evaluate segmentation performance.

4.1. Comparisons with State-of-the-art

Quantitative results: In Tab. 2, we compare our approach with SOTA methods including HeMIS [22], U-HVED [14], D2-Net [53], ACN [52], RFNet [13] and mmFormer [56] on BRATS2018. HeMIS, U-HVED, and D2-Net learn a biased mapping among available modalities and hence perform poorly compared to ACN which co-trains with the full modality of all samples. Recent shared latent space models like mmFormer and RFNet perform comparably to ACN. They either focus on learning inter-modal correlations or tumor region-aware fused representations. Our method utilizes full modality data from only 50% samples, and yet outperforms these approaches. We thus excel in efficient utilization of full modality data. In comparison with the second-best approach in WT, TC, and ET, our average DSC shows improvements of 0.89% (over mmFormer), 1.96% (over ACN), and 1.68% (over ACN), respectively. Although ACN pursues a KD-driven approach to achieve the partial-to-full modality mapping, it ends up building a combinatorial number of models dedicated to each subset. This leads to a highly ineffective solution which is also based on the impractical scenario that all training samples contain full modality data. In our framework, we mimic this distillation learning even in a shared latent model through efficient application of meta-learning and adversarial training. It can be seen from Tab. 2 that our method surpasses all other approaches in 39 out of 45 multi-modal combinations across the three tumor regions despite being trained with only 50% full modality samples. Our results are statistically significantly better (t-test, $p < 0.05$) than HeMIS [22], U-HVED [14], D2-Net [53]. Other methods (RFNet [13],

M																	
	FLAIR	○	○	○	•	○	○	•	○	•	•	•	•	•	○	•	Avg ± std, p-value (10^{-2})
	T1	○	○	•	○	○	•	•	•	○	○	•	•	○	•	•	
	T1c	○	•	○	○	•	•	○	○	○	•	•	○	•	•	•	
	T2	•	○	○	○	•	○	○	•	•	○	○	•	•	•	•	
WT	HeMIS[22]	79.85	60.32	55.76	66.20	81.63	65.39	75.41	81.70	82.56	76.25	79.82	84.58	86.27	82.74	85.06	76.23±9.66, 0.05*
	U-HVED[14]	81.06	58.74	52.37	82.65	80.88	66.21	83.70	82.83	86.44	84.92	86.33	87.56	87.84	83.47	88.25	79.55±11.14, 2.10*
	D2-Net[53]	76.58	43.79	19.43	85.06	84.62	65.37	86.18	82.56	86.35	87.94	87.31	88.59	89.12	83.78	88.94	77.04±19.95, 6.64
	ACN[52]	85.24	79.16	78.65	86.72	85.87	79.27	86.33	85.21	86.69	87.54	86.92	88.22	87.51	86.38	89.14	85.25±3.38, 20.43
	RFNet[13]	84.92	73.41	72.57	87.93	86.22	78.31	89.43	**86.81**	89.98	89.23	89.80	90.22	90.16	86.95	90.32	85.75±6.03, 48.41
	mmFormer[56]	84.73	76.10	75.39	**88.53**	86.75	79.24	89.57	86.61	90.05	89.69	89.64	90.11	90.20	87.11	90.09	86.25±5.16, 62.41
	Ours	**86.52**	**79.23**	**78.66**	87.45	**86.77**	**79.60**	**89.94**	86.71	**90.82**	**90.13**	**90.38**	**90.74**	**90.63**	**88.09**	**91.26**	**87.12±4.43**
TC	HeMIS[22]	49.63	53.75	24.80	32.91	70.28	64.29	45.62	54.36	54.93	69.40	72.57	62.38	75.51	73.94	74.18	58.57±15.50, 0.01*
	U-HVED[14]	56.62	64.50	36.77	54.38	74.46	65.29	59.03	58.66	62.57	73.14	75.85	63.72	73.52	76.81	72.96	64.55±10.72, 0.02*
	D2-Net[53]	59.87	64.29	20.32	50.84	81.06	77.96	62.54	64.18	61.70	82.45	79.38	67.52	81.47	80.23	80.94	67.65±16.55, 2.02*
	ACN[52]	67.24	84.35	70.49	67.38	84.70	83.92	70.61	**73.58**	70.66	82.17	84.35	67.08	81.94	84.32	84.73	77.16±7.56, 47.26
	RFNet[13]	67.72	78.87	64.39	67.85	83.04	80.84	72.80	71.65	73.32	83.76	84.09	74.89	84.26	82.98	84.40	76.99±7.00, 41.76
	mmFormer[56]	65.92	77.50	62.94	66.10	80.58	79.35	72.31	69.89	71.39	79.72	81.53	73.30	80.68	80.56	81.62	74.89±6.46, 10.09
	Ours	**68.12**	**84.57**	**71.24**	**68.75**	**85.67**	**84.39**	**73.48**	72.90	**73.71**	**84.97**	**85.43**	**75.62**	**84.75**	**86.56**	**86.77**	**79.12±7.18**
ET	HeMIS[22]	22.47	56.20	7.89	9.64	64.07	65.66	17.73	26.95	27.42	65.83	70.35	30.18	68.97	69.52	73.80	45.11±24.97, 3.23*
	U-HVED[14]	27.82	61.24	11.06	22.35	68.93	65.79	24.57	24.46	35.80	69.31	71.42	32.14	70.66	69.98	71.20	48.44±22.98, 6.41
	D2-Net[53]	22.83	69.52	15.34	12.96	70.45	71.38	14.06	19.32	17.79	69.25	68.31	23.66	67.14	68.56	67.72	45.22±26.52, 4.07*
	ACN[52]	43.26	**78.57**	40.89	42.14	74.95	75.88	42.73	**47.80**	44.39	76.72	76.33	41.61	75.54	75.27	76.79	60.85±17.12, 78.65
	RFNet[13]	40.62	69.73	37.62	38.08	75.42	71.55	45.67	43.44	45.36	75.18	**76.52**	47.14	76.75	75.26	76.71	59.67±16.85, 64.25
	mmFormer[56]	39.65	66.23	37.08	38.72	68.70	67.84	45.15	42.61	43.69	68.42	68.36	45.33	68.45	69.81	68.16	55.88±13.86, 24.25
	Ours	**44.87**	78.09	**41.12**	**43.94**	**77.16**	**77.58**	**45.81**	46.25	**48.63**	**77.29**	76.04	**48.22**	**77.92**	**76.71**	**78.30**	**62.53±16.53**

Table 2: Comparison with state-of-the-art (DSC %) on segmentation of nested tumor regions (WT, TC, ET) for the different combinations of available modalities on BRATS2018. Our approach trains with 50% full modality samples while others use 100%. The best and second best scores are **bolded** and underlined, respectively. Modalities present are denoted by •, the missing ones by ○. Statistically significant results with p-values ≤ 0.05 are denoted by $*$.

mmFormer [56], ACN [52]) require full modality input for all samples, whereas ours does not. However, for a fair comparison, we further trained [13, 56, 52] in a 50% full modality setting, identical to ours (Tab. 3). It can be observed that our method outperforms the second-best approach by 11.78%, 12.93%, and 9.72% in DSC on the WT, TC, and ET regions, respectively, on BRATS2018, clearly achieving SOTA performance. Evaluation via HD95 metric on BRATS2018 can be found in the supplementary (Sec. 9). Comparison with three SOTA methods on BRATS2020 are presented in Tab. 5; Compared to the second-best approach, the average DSC of the three tumor areas is boosted by 1.78%, 2.84%, and 3.13%, respectively. Detailed experiments on BRATS2020 and BRATS2019 have been provided in the supplementary (Sec. 10).

Qualitative results: In Fig. 5 we visualize the segmentation masks predicted by U-HVED, RFNet, and ours from four combinations of modalities. Unlike others, our segmentations do not degrade sharply as additional modalities are dropped during the inference phase. Even with single T2 or T1c+T2 modalities, we achieve decent segmentation.

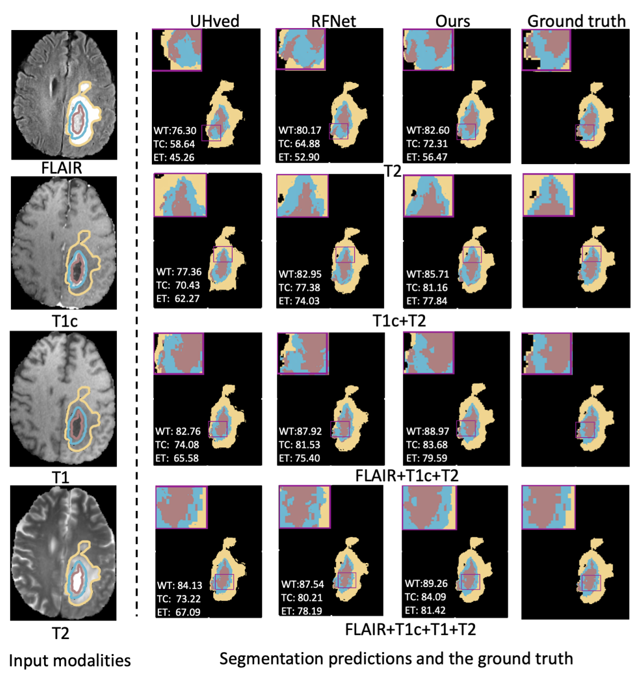


Figure 5: Qualitative comparison. Column 1: four MRI modalities. Col 2-4: segmentation maps from three methods for different combinations of modalities. Col 5: Ground truth. Our method is able to better capture gaps/islands (rows 1,3) and boundaries (row 4) in TC segmentations.

4.2. Ablation Studies

Effectiveness of adversarial and meta-learning. We perform several ablations to evaluate and justify the contribution of each proposed module in our architecture. First, we

Methods	Average DSC (%), p-value (10^{-2})		
	WT	TC	ET
ACN[52]	65.81, 0.01*	57.66, 0.01*	47.36, 1.63*
RFNet[13]	73.96, 0.01*	62.37, 0.01*	50.24, 3.80*
mmFormer[56]	75.34, 0.01*	66.19, 0.01*	52.81, 7.64
Ours	**87.12**	**79.12**	**62.53**

Table 3: Comparison (DSC%, p-value) on BRATS2018 with 50% full modality

Methods	Average DSC (%)		
	WT	TC	ET
mDrop	81.67	72.41	52.06
+ GAN	84.49	75.38	55.64
+ MetaL	85.96	77.23	59.85
+ GAN + MetaL	**87.12**	**79.12**	**62.53**

Table 4: Ablation study demonstrating effectiveness of major components.

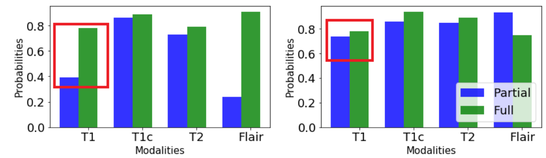
Methods	Average DSC (%), p-value (10^{-2})		
	WT	TC	ET
HeMIS[22]	77.29, 0.01*	67.12, 0.13*	49.64, 4.25*
U-HVED[14]	82.65, 1.49*	69.08, 0.22*	51.53, 5.84
RFNet[13]	86.96, 23.71	78.79, 27.86	62.14, 59.95
Ours	**88.74**	**81.63**	**65.27**

Table 5: Comparison (DSC%, p-value) on BRATS2020.

remove both the Adversarial and the Meta-training strategies to perform segmentation from only fusion of available modalities. We thus formulate a baseline, *mDrop*, where we include our feature-aggregation block to generate a fused representation from available modalities. *mDrop* solely learns the intra-model relations through transformer encoders and inter-modal dependencies through channel-weighted fusion. The average DSC of our model outperforms *mDrop* by 5.45%, 6.71%, and 10.47% in the three tumor regions (Tab. 4). Hence it is evident that solely the modality-agnostic representations obtained from fusion of available modalities cannot generate accurate segmentations. This necessitates feature enrichment to improve the quality of the fused representation. We develop two variants through gradual introduction of our discriminator and meta-learning strategies as enrichment techniques. Both variants surpass mDrop considerably (Tab. 4). Meta-learning (MetaL) proved to be better since we built the heterogeneous task distribution with modality combinations (to reduce bias) and also explicitly adapted to the full modality feature space efficiently. Finally, we arrive at an end-to-end meta-learning framework that also benefits from auxiliary supervision provided by the adversarial discriminator.

Evaluation of enhanced representations. To evaluate the quality of enhanced representations, we designed a simple experiment. We first extract the bottleneck fused representations $\mathbf{F}^B_{fused}$ of 50 test subjects for both scenarios of full modality (where all are present) and partial modality (where only T1c, T2 are present). This was done for both *mDrop* as well as our approach. The fused representations were fed into a classifier trained to predict the probabilities of a modality's presence. The average probabilities obtained from our method attain comparable distributions across partial and full modality settings (Fig. 6), depicting the desired enhancement of $\mathbf{F}^B_{fused}$. However, for *mDrop*, probabilities of T1c and T2 being present are much higher than T1, and FLAIR in the missing scenario. Hence the relevant information from the latter two modalities is being lost. The red boxes depicts how our probability for predicting T1 is considerably higher than *mDrop* even in T1-missing scenario.

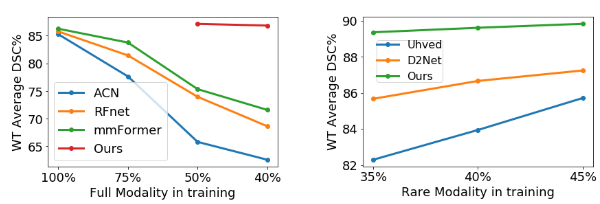
Robustness to full modality setting. Due to the meta-learning strategy incorporated while training on hybrid data, we hypothesize that our network is robust to the ratio of full modality samples used in training. We compare against



(a) *mDrop* (b) Ours

Figure 6: Comparison of baseline and enhanced features

ACN, RFNet, and mmFormer by varying the full modality count from 100% to 40% (Fig. 7a). In order to retain sufficient samples for each combination task in meta-training, we assume that at least 50% of the subjects have partial modalities. Hence we show our results only on 50% and 40% proportions of full modality data. The fact that even with 50% full modality samples, we match the evaluation scores of SOTA at 100% setting is noteworthy. A sharp degradation can be noticed in the average WT DSC of SOTA once the number of full modality data decreases. On the other hand, our method shows only a minor drop of 0.29%. This is due to ACN being heavily dependent on the full modality for knowledge distillation. RFNet and mmFormer require full modality data as input to the network. They under-fit since their overall sample count decreases. Our method efficiently utilizes even limited samples of full modality data for feature adaptation in meta-testing. Owing to the above reasons, our approach is resilient to change in full modality proportion. Results for other tumor regions are provided in supplementary (Sec. 8).



(a) Ablation results for varying % of full modality in training. (b) Ablation results for varying % of FLAIR in training.

Figure 7: Ablation studies.

Bias to presence of a specific modality. Our model is robust to the scenario when a modality appears rarely during training. Tab. 6 demonstrates that when only 35% of FLAIR is considered for training, our method consistently outper-

M										Avg
	FLAIR	●	●	●	●	●	●	●	●	
	T1	○	●	○	○	●	●	○	●	
	T1c	○	○	○	●	●	○	●	●	
	T2	○	○	●	○	○	●	●	●	
WT	U-HVED	76.61	78.94	83.98	80.47	82.76	84.39	85.08	86.25	82.30
	D2-Net	82.44	83.79	85.16	85.55	85.18	87.63	88.02	87.70	<u>85.68</u>
	Ours	**86.53**	**88.77**	**90.05**	**89.41**	**89.49**	**90.13**	**90.07**	**90.55**	**89.37**
TC	U-HVED	50.23	55.46	57.35	71.85	74.31	59.02	72.19	71.40	63.97
	D2-Net	47.19	59.98	59.65	81.23	77.44	64.86	80.11	79.70	<u>68.77</u>
	Ours	**67.26**	**72.28**	**72.64**	**84.03**	**84.87**	**74.08**	**83.86**	**85.91**	**78.11**
ET	U-HVED	17.58	20.39	32.17	67.75	70.11	28.94	69.60	70.09	<u>47.07</u>
	D2-Net	9.37	11.78	14.09	68.14	67.58	20.22	66.31	66.80	40.53
	Ours	**42.58**	**44.27**	**47.14**	**76.12**	**74.95**	**46.83**	**76.91**	**77.61**	**60.80**

Table 6: Results for rare occurrence of FLAIR in training.

forms U-HVED and D2-Net in all the 8 inference scenarios involving FLAIR. Our average DSC (89.37, 78.11, and 60.80) for (WT, TC, and ET) are significantly higher than the second-best method (85.68, 68.77, and 47.07). We attribute this improvement to meta-learning which precludes the model from learning a biased mapping among available modalities by aligning the shared representations to full modality representations. Further experiments in Fig. 7b demonstrate that the performance of U-HVED and D2-Net are highly sensitive to the availability of a particular modality while our approach is impervious to this. On increasing FLAIR from 35% to 45%, our WT gain (+0.47%) is much lower than U-HVED (+3.42%) or D2-Net (+1.57%). Experiments with another modality (T1c) are provided in supplementary (Sec. 11).

Robustness to backbone variants. The proposed meta and adversarial training strategies are robust to any employed backbone including 3DUnet [12], nnUnet [24] and AttentionUnet [38]. Comparisons are provided in supplementary (Sec. 7).

Ablation on aggregation block. We design 3 baseline aggregation modules to highlight the effectiveness of our fusion strategy. Architectural details are provided in supplementary (Sec. 13).

5. Conclusion

We present a novel training strategy to address the problem of missing modalities in brain tumor segmentation under limited full modality supervision. We adopt meta-learning and formulate modality combinations as separate meta-tasks to mitigate the bias towards modalities rarely encountered in training. We distill discriminative features from full modality data in the meta-testing phase, thereby discarding the impractical omnipresence of full modalities for all samples. This mapping is further co-supervised by novel adversarial learning in latent space, that guarantees the generation of superior modality-agnostic representations. In the future we will validate our method on other downstream tasks such as radiogenomics classification [44] and treatment response prediction [41].

6. Acknowledgements

Reported research was partially supported by NIH 1R21CA258493-01A1 and NSF CCF-2144901. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] Antreas Antoniou, Harrison Edwards, and Amos Storkey. How to train your MAML. In *ICLR*, 2019. 3

[2] Reza Azad, Nika Khosravi, and Dorit Merhof. SMU-net: Style matching U-Net for brain tumor segmentation with missing modalities. In *MIDL*, 2022. 1, 2, 3

[3] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific data*, 2017. 1

[4] Mahsa Baktashmotlagh, Mehrtash T Harandi, Brian C Lovell, and Mathieu Salzmann. Unsupervised domain adaptation by domain invariant projection. In *ICCV*, 2013. 3

[5] Stefan Bauer, Roland Wiest, Lutz-P Nolte, and Mauricio Reyes. A survey of MRI-based medical image analysis for brain tumor studies. *Physics in Medicine & Biology*, 2013. 1

[6] Moinak Bhattacharya, Shubham Jain, and Prateek Prasanna. Radiotransformer: A cascaded global-focal transformer for visual attention-guided disease classification. In *ECCV*, 2022. 3

[7] Agisilaos Chartsias, Thomas Joyce, Mario Valerio Giuffrida, and Sotirios A Tsaftaris. Multimodal MR synthesis via modality-invariant latent representation. *TMI*, 2017. 2, 3

[8] Cheng Chen, Qi Dou, Yueming Jin, Hao Chen, Jing Qin, and Pheng-Ann Heng. Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion. In *MICCAI*, 2019. 2, 3

[9] Jiayi Chen and Aidong Zhang. HetMAML: Task-heterogeneous model-agnostic meta-learning for few-shot learning across modalities. In *CIKM*, 2021. 3, 4

[10] Shengcong Chen, Changxing Ding, and Minfeng Liu. Dual-force convolutional neural networks for accurate brain tumor segmentation. *Pattern Recognition*, 2019. 1, 3

[11] Jae Won Cho, Dong-Jin Kim, Jinsoo Choi, Yunjae Jung, and In So Kweon. Dealing with missing modalities in the visual question answer-difference prediction task through knowledge distillation. In *CVPR*, 2021. 3

[12] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *MICCAI*, 2016. 9

[13] Yuhang Ding, Xin Yu, and Yi Yang. RFNet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation. In *ICCV*, 2021. 2, 3, 6, 7, 8

[14] Reuben Dorent, Samuel Joutard, Marc Modat, Sébastien Ourselin, and Tom Vercauteren. Hetero-modal variational

encoder-decoder for joint modality completion and segmentation. In *MICCAI*, 2019. 2, 3, 6, 7, 8

[15] Changde Du, Changying Du, Hao Wang, Jinpeng Li, Wei-Long Zheng, Bao-Liang Lu, and Huiguang He. Semi-supervised deep generative modelling of incomplete multi-modality emotional data. In *ACM Multimedia*, 2018. 3

[16] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, 2017. 2, 3

[17] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *ICML*, 2015. 3

[18] Martin J Graves and Donald G Mitchell. Body MRI artifacts in clinical practice: a physicist's and radiologist's perspective. *Journal of Magnetic Resonance Imaging*, 2013. 1

[19] Arthur Gretton, Alex Smola, Jiayuan Huang, Marcel Schmittfull, Karsten Borgwardt, and Bernhard Schölkopf. Covariate shift by kernel mean matching. *Dataset shift in machine learning*, 2009. 3

[20] Mohammad Hamghalam, Alejandro F Frangi, Baiying Lei, and Amber L Simpson. Modality completion via gaussian process prior variational autoencoders for multi-modal glioma segmentation. In *MICCAI*, 2021. 2, 3

[21] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *MedIA*, 2017. 1, 3

[22] Mohammad Havaei, Nicolas Guizard, Nicolas Chapados, and Yoshua Bengio. HeMIS: Hetero-modal image segmentation. In *MICCAI*, 2016. 2, 3, 6, 7, 8

[23] Minhao Hu, Matthis Maillard, Ya Zhang, Tommaso Ciceri, Giammarco La Barbera, Isabelle Bloch, and Pietro Gori. Knowledge distillation from multi-modal to mono-modal segmentation networks. In *MICCAI*, 2020. 1, 2, 3

[24] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 2021. 9

[25] Mobarakol Islam, Navodini Wijethilake, and Hongliang Ren. Glioblastoma multiforme prognosis: MRI missing modality generation, segmentation and radiogenomic survival prediction. *Computerized Medical Imaging and Graphics*, 2021. 1, 2, 3

[26] Aishik Konwer, Xuan Xu, Joseph Bae, Chao Chen, and Prateek Prasanna. Temporal context matters: Enhancing single image prediction with disease progression representations. In *CVPR*, 2022. 3

[27] Gueorgi Kossinets. Effects of missing data in social networks. *Social networks*, 2006. 3

[28] Kenneth Lau, Jonas Adler, and Jens Sjölund. A unified representation network for segmentation with missing modalities. *arXiv preprint arXiv:1908.06683*, 2019. 2, 3

[29] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy Hospedales. Learning to generalize: Meta-learning for domain generalization. In *AAAI*, 2018. 3

[30] Han Liu, Yubo Fan, Hao Li, Jiacheng Wang, Dewei Hu, Can Cui, Ho Hin Lee, Huahong Zhang, and Ipek Oguz. Moddrop++: A dynamic filter network with intra-subject co-training for multiple sclerosis lesion segmentation with missing modalities. In *MICCAI*, 2022. 2, 3

[31] Mengmeng Ma, Jian Ren, Long Zhao, Sergey Tulyakov, Cathy Wu, and Xi Peng. SMIL: Multimodal learning with severely missing modality. In *AAAI*, 2021. 2, 3, 4

[32] Yao Ma, Shilin Zhao, Weixiao Wang, Yaoman Li, and Irwin King. Multimodality in meta-learning: A comprehensive survey. *Knowledge-Based Systems*, 2022. 3

[33] J MacQueen. Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability*, 1967. 2

[34] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *TMI*, 2014. 6

[35] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3DV*, 2016. 6

[36] Andriy Myronenko. 3D MRI brain tumor segmentation using autoencoder regularization. In *MICCAI Brainlesion Workshop*, 2018. 1, 3

[37] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018. 3

[38] Ozan Oktay, Jo Schlemper, Loic Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning where to look for the pancreas. In *MIDL*, 2018. 9

[39] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NeurIPS Workshop*, 2017. 6

[40] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 1901. 2

[41] Prateek Prasanna, Jay Patel, Sasan Partovi, Anant Madabhushi, and Pallavi Tiwari. Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-parametric mr imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings. *European radiology*, 27:4188–4197, 2017. 9

[42] Anmol Sharma and Ghassan Hamarneh. Missing MRI pulse sequence synthesis using multi-modal generative adversarial network. *TMI*, 2019. 1, 2, 3

[43] Yan Shen and Mingchen Gao. Brain tumor segmentation on MRI with missing modalities. In *IPMI*, 2019. 1, 3

[44] Gagandeep Singh, Sunil Manjila, Nicole Sakla, Alan True, Amr H Wardeh, Niha Beig, Anatoliy Vaysberg, John Matthews, Prateek Prasanna, and Vadim Spektor. Radiomics and radiogenomics in gliomas: a contemporary update. *British journal of cancer*, 125(5):641–657, 2021. 9

[45] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *NeurIPS*, 2017. 3

[46] Baochen Sun and Kate Saenko. Deep CORAL: Correlation alignment for deep domain adaptation. In *ECCV*, 2016. 3

[47] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In *CVPR*, 2019. 3

[48] Qiuling Suo, Weida Zhong, Fenglong Ma, Ye Yuan, Jing Gao, and Aidong Zhang. Metric learning on healthcare data with incomplete modalities. In *IJCAI*, 2019. 3

[49] Yucheng Tang, Dong Yang, Wenqi Li, Holger R Roth, Bennett Landman, Daguang Xu, Vishwesh Nath, and Ali Hatamizadeh. Self-supervised pre-training of swin transformers for 3D medical image analysis. In *CVPR*, 2022. 1, 3, 6

[50] Saverio Vadacchino, Raghav Mehta, Nazanin Mohammadi Sepahvand, Brennan Nichyporuk, James J Clark, and Tal Arbel. HAD-Net: A hierarchical adversarial knowledge distillation network for improved enhanced tumour segmentation without post-contrast images. In *MIDL*, 2021. 1, 2, 3

[51] Wenxuan Wang, Chen Chen, Meng Ding, Hong Yu, Sen Zha, and Jiangyun Li. Transbts: Multimodal brain tumor segmentation using transformer. In *MICCAI*, 2021. 1, 3

[52] Yixin Wang, Yang Zhang, Yang Liu, Zihao Lin, Jiang Tian, Cheng Zhong, Zhongchao Shi, Jianping Fan, and Zhiqiang He. ACN: Adversarial co-training network for brain tumor segmentation with missing modalities. In *MICCAI*, 2021. 1, 2, 3, 6, 7, 8

[53] Qiushi Yang, Xiaoqing Guo, Zhen Chen, Peter YM Woo, and Yixuan Yuan. D2-net: Dual disentanglement network for brain tumor segmentation with missing modalities. *TMI*, 2022. 2, 3, 6, 7

[54] Biting Yu, Luping Zhou, Lei Wang, Jurgen Fripp, and Pierrick Bourgeat. 3D cGAN based cross-modality MR image synthesis for brain tumor segmentation. In *ISBI*, 2018. 1, 2, 3

[55] Ziqi Yu, Yuting Zhai, Xiaoyang Han, Tingying Peng, and Xiao-Yong Zhang. MouseGAN: GAN-based multiple MRI modalities synthesis and segmentation for mouse brain structures. In *MICCAI*, 2021. 1, 2, 3

[56] Yao Zhang, Nanjun He, Jiawei Yang, Yuexiang Li, Dong Wei, Yawen Huang, Yang Zhang, Zhiqiang He, and Yefeng Zheng. mmFormer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation. In *MICCAI*, 2022. 2, 3, 6, 7, 8

[57] Zechen Zhao, Heran Yang, and Jian Sun. Modality-adaptive feature interaction for brain tumor segmentation with missing modalities. In *MICCAI*, 2022. 2, 3

[58] Chenhong Zhou, Changxing Ding, Xinchao Wang, Zhentai Lu, and Dacheng Tao. One-pass multi-task networks with cross-task guided attention for brain tumor segmentation. *TIP*, 2020. 1, 3

[59] Tongxue Zhou, Stéphane Canu, Pierre Vera, and Su Ruan. Conditional generator and multi-sourcecorrelation guided brain tumor segmentation with missing MR modalities. *arXiv preprint arXiv:2105.13013*, 2021. 1, 2, 3

[60] Tongxue Zhou, Stéphane Canu, Pierre Vera, and Su Ruan. Latent correlation representation learning for brain tumor segmentation with missing MRI modalities. *TIP*, 2021. 2, 3